JMB

Microbial Genome Analyses: Global Comparisons of Transport Capabilities Based on Phylogenies, Bioenergetics and Substrate Specificities

Ian T. Paulsen, Marek K. Sliwinski and Milton H. Saier, Jr*

Department of Biology
University of California at San
Diego, La Jolla
CA 92093-0116, USA

We have conducted genome sequence analyses of seven prokaryotic microorganisms for which completely sequenced genomes are available (*Escherichia coli, Haemophilus influenzae, Helicobacter pylori, Bacillus subtilis, Mycoplasma genitalium, Synechocystis* PCC6803 and *Methanococcus jannaschii*). We report the distribution of encoded known and putative polytopic cytoplasmic membrane transport proteins within these genomes. Transport systems for each organism were classified according to (1) putative membrane topology, (2) protein family, (3) bioenergetics and (4) substrate specificities. The overall transport capabilities of each organism were thereby estimated. Probable function was assigned to greater than 90% of the putative transport proteins identified. The results show the following: (1) Numbers of transport systems in eubacteria are approximately proportional to genome size and correspond to 9.7 to 10.8% of the total encoded genes except for *H. pylori* (5.4%), *Synechocystis* (4.7%) and *M. jannaschii* (3.5%) which exhibit substantially lower proportions. (2) The distribution of topological types is similar in all seven organisms. (3) Transport systems belonging to 67 families were identified within the genomes of these organisms, and about half of these families are also found in eukaryotes. (4) 12% of these families are found exclusively in Gram-negative bacteria, but none is found exclusively in Gram-positive bacteria, cyanobacteria or archaea. (5) Two superfamilies, the ATP-binding cassette (ABC) and major facilitator (MF) superfamilies account for nearly 50% of all transporters in each organism, but the relative representation of these two transporter types varies over a tenfold range, depending on the organism. (6) Secondary, pmf-dependent carriers are 1.5 to threefold more prevalent than primary ATP-dependent carriers in *E. coli*, *H. influenzae, H. pylori* and *B. subtilis* while primary carriers are about twofold more prevalent in *M. genitalium* and *Synechocystis*. *M. jannaschii* exhibits a slight preference for secondary carriers. (7) Bioenergetics of transport generally correlate with the primary forms of energy generated *via* available metabolic pathways but ecological niche and substrate availability may also be determining factors. (8) All organisms display a similar range of transport specificities with quantitative differences presumably reflective of disparate ecological niches. (9) *M. jannaschii* and *Synechocystis* have a two to threefold increased proportion of transporters for inorganic ions with a concomitant decrease in transporters for organic compounds. (10) 6 to 18% of all transporters in these bacteria probably function as drug export systems showing that these systems are prevalent in non-pathogenic as well as pathogenic organisms. (11) All seven prokaryotes examined encode proteins homologous to known channel proteins, but none of the channel types identified occurs in all of these

*Corresponding author

organisms. (12) The phosphoenolpyruvate:sugar phosphotransferase system is prevalent in the large genome organisms, *E. coli* and *B. subtilis*, and is present in the small genome organisms, *H. influenzae* and *M. genitalium*, but is totally lacking in *H. pylori*, *Synechocystis* and *M. jannaschii*. Details of the information summarized in this article are available on our web sites, and this information will be periodically updated and corrected as new sequence and biochemical data become available.

© 1998 Academic Press Limited

Introduction

The last three years have seen impressive progress in the sequencing of entire genomes of both prokaryotic and eukaryotic free-living organisms. As of October, 1997, the complete genomic sequences for *Haemophilus influenzae* (Fleischmann *et al.*, 1995), *Mycoplasma genitalium* (Fraser *et al.*, 1995), *Methanococcus jannaschii* (Bult *et al.*, 1996), *Synechocystis* PCC6803 (Kaneko *et al.*, 1996), *Mycoplasma pneumoniae* (Himmelreich *et al.*, 1996), *Saccharomyces cerevisiae* (Goffeau *et al.*, 1997), *Escherichia coli* (Blattner *et al.*, 1997), *Helicobacter pylori* (Tomb *et al.*, 1997) and *Bacillus subtilis* (Kunst *et al.*, 1997) were publicly available. The wealth of data from these and forthcoming genome sequencing projects are likely to promote a revolution in the biological sciences. However, the impact resulting from the availability of these data is not likely to be felt until systematic analyses have been conducted.

Tremendous effort has been devoted to the characterization and classification of enzymes comprising key metabolic pathways in microorganisms (Karp, *et al.*, 1996; Karp & Paley, 1996; Selkov *et al.*, 1996). However, the utilization of any exogenous substrate depends on the activities of cytoplasmic membrane transport systems that allow entry of the potential metabolite into the cell cytoplasm. Further, excretion of metabolic end products depends on such transport systems, and the pathways for entry of a particular substrate may differ from those for exit (Krämer, 1994). The characterization and classification of transport proteins is unfortunately far less advanced than those of metabolic enzymes.

Transport proteins are known to catalyze transmembrane solute translocation by several distinct mechanisms (Mitchell, 1967a,b; Saier, 1998). Thus, solute-non-specific and solute-specific channels as well as highly stereospecific carriers have been studied extensively (Fischer *et al.*, 1995; Konings *et al.*, 1996; Lee, 1996). Moreover, a variety of energy coupling mechanisms are known to drive the active uptake and/or extrusion of specific compounds (Fath & Kolter, 1993; Krämer, 1994; Konings *et al.*, 1996; Paulsen *et al.*, 1996a,b). The energy sources most commonly used include chemical energy in the form of ATP or phosphoe-

nolpyruvate (PEP) and chemiosmotic energy in the form of a sodium or proton electrochemical gradient, also known as the sodium or proton motive force (smf or pmf, respectively (Maloney, 1990, 1992). However, other forms of energy are sometimes utilized to drive transport (Dimroth, 1997; Saier, 1998).

Transport proteins (porters) usually exhibit specificity for a restricted range of substrates (Clark & Amara, 1993; Poolman *et al.*, 1996). Distinct families of porters have been described on the basis of sequence similarities (Saier & Reizer, 1991). Within each family, phylogenetic analyses have revealed that substrate specificity frequently correlates with phylogeny (Saier, 1994, 1996, 1998). This observation implies that substrate specificity is often a well conserved evolutionary trait. Phylogenetic analyses therefore provide a reliable basis for functional analysis.

Cytoplasmic membrane transport proteins typically consist of multiple membrane spanning α-helical segments (TMSs) connected by loop regions of various sizes (Kaback, 1986). Based in part on hydropathy analyses, and confirmed experimentally for several such proteins, a surprising number of these proteins are believed to exhibit either six or 12 of these transmembrane segments (Yan & Maloney, 1990, 1992; Saier, 1994; Dean & Allikmets, 1995; Kuan *et al.*, 1995; Pourcher *et al.*, 1996). While sequence comparisons indicate that a putative six TMS topological unit has evolved independently many times, the architectural basis for this topological uniformity is not present understood (Henderson & Maiden, 1990; Griffith *et al.*, 1992; Maloney, 1990, 1992; Saier, 1998). Models suggesting specific three-dimensional topologies of these proteins have been proposed, but three-dimensional structural data are not yet available (Kaback, *et al.*, 1997; Yan & Maloney, 1993, 1995; Goswitz & Brooker, 1995; Le & Saier, 1996).

To date, few systematic and comprehensive analyses of the transport capabilities of microorganisms have been reported (Clayton *et al.*, 1997). In order to correct this deficiency, we have initiated a study to identify and characterize all existing cytoplasmic membrane porters in those prokaryotic microorganisms for which extensive genome sequence data are available. In this capacity we

have systematically analyzed the complete genomic sequences of the eight prokaryotes cited above. Comparable analyses of the complete genome of the eukaryote *Saccharomyces cerevisiae* are currently underway and will be the subject of a separate report (I.T.P., M.K.S. & M.H.S., Jr, unpublished). These analyses have allowed us to identify members of more than 67 families of transport systems encoded within these genomes (see Table 2 and our web sites for the detailed analyses).

Here, we present the categorization of these proteins according to their energy coupling mechanisms, their families, their substrate specificities, and their transmembrane topologies. The reported analyses allow an overall evaluation of the transport capabilities of these organisms. They also allow estimation of the number of transport protein families yet to be identified and characterized in bacteria. Some of the unexpected findings are interpreted and discussed.

Results

Overall approach for the identification of membrane transport systems

To identify all of the known and putative cytoplasmic membrane transporters encoded within the genome of each organism, we utilized an approach based on the fact that bacterial cytoplasmic membrane transport systems characteristically include at least one membrane protein which contains hydrophobic regions predicted to form multiple, transmembrane, hydrophobic, α-helical segments (TMSs; Konings *et al.*, 1996; Saier, 1994, 1996). We performed systematic hydropathy analyses on all of the predicted protein products encoded within the genome of each organism analyzed and predicted the likely number of TMSs in each protein (see Table 1 below). Proteins with one or more multiple hydrophobic TMS(s) were then screened against the protein databases (see

Materials and Methods) in order to identify functionally characterized homologs that would provide an indication of the function of the test protein. Transport system families identified were subsequently classified on the basis of sequence similarities of protein members, bioenergetics, and substrate specificities.

Hydropathy analyses

Systematic hydropathy analyses of all of the recognized proteins encoded within the complete genomes of the organisms studied enabled prediction of membrane topologies for these proteins. An overall comparison of the relative abundances of membrane proteins of differing predicted topologies for each organism is provided in Table 1. As can be seen, there is a similar distribution of protein topological types in all of the organisms analyzed.

Overall, about 70% (65 to 77% for the different organisms examined; Table 1) of the encoded proteins are predicted to be soluble or peripheral membrane proteins while only about 30% are integral membrane proteins. The organisms with the largest genomes exhibit the largest proportion of putative integral membrane proteins while the archaeon, *M. jannaschii*, exhibits the smallest proportion. 11 to 19% of the proteins, or about half of the integral membrane proteins, possess only one putative TMS. Most of these proteins probably do not serve a primary transport function although some of them clearly serve in auxiliary transport capacities (Dinh *et al.*, 1994; Paulsen *et al.*, 1997; see above). Many of the one TMS proteins are likely to possess cleavable N-terminal signal sequences that target the protein to the periplasm or outer membrane (Saier *et al.*, 1989). About 4 to 7% of the proteins exhibit two to three putative TMSs. Some of these proteins undoubtedly serve primary transport functions in the cytoplasmic membrane, as is true for example of the MscL mechanosensitive ion

Table 1. Comparison of relative abundances of proteins of differing predicted membrane topologies for eight bacteria

Organism (organismal classification)	Number of predicted TMS					
	0	1	2–3	4–6	7–9	⩾10
Escherichia coli (Gram-negative bacterium)	2861 (66.8%)	655 (15.3 %)	220 (5.1%)	211 (4.9%)	153 (3.6%)	182 (4.3%)
Haemophilus influenzae (Gram-negative bacterium)	1204 (72.0%)	223 (13.3%)	67 (4.0%)	85 (5.1%)	53 (3.2%)	41 (2.4%)
Helicobacter pylori (Gram-negative bacterium)	1121 (71.1%)	239 (15.2%)	86 (5.5%)	72 (4.6%)	27 (1.7%)	32 (2.0%)
Bacillus subtilis (Gram-positive bacterium)	2686 (66.8%)	546 (13.6%)	266 (6.6%)	232 (5.8%)	148 (3.7%)	143 (3.6%)
Mycoplasma genitalium (Gram-positive bacterium)	318 (68.7%)	72 (15.6%)	24 (5.1%)	24 (5.1%)	17 (3.7%)	8 (1.7%)
Mycoplasma pneumoniae (Gram-positive bacterium)	474 (70.0%)	104 (15.4%)	34 (5.0%)	37 (5.5%)	18 (2.7%)	10 (1.5%)
Synechocystis (Cyanobacterium)	2058 (65.0%)	593 (18.7%)	214 (6.8%)	160 (5.1%)	84 (2.7%)	58 (1.8%)
Methanococcus jannaschii (Archaeon)	1341 (77.3%)	194 (11.2%)	76 (4.4%)	79 (4.6%)	26 (1.5%)	19 (1.1%)

Table 2. Transport protein families identified by bacterial genome analysis

Name of family	Abbreviation	TC no.[a]	Example[b]	Typical substrates	Distribution[c]	References
I. Channel proteins						
Major intrinsic protein	MIP	1.1	GlpF Eco	Glycerol, water	BE	Park & Saier (1996)
Large conductance mechano-sensitive ion channel	MscL	1.3	MscL Eco	Ions	B	Sukarev et al. (1996)
Voltage-sensitive ion channel	VIC	1.5	Kch Eco	Potassium	BAF	Alexander & Peters (1997)
Chloride channel	ClC	1.10	GefI See	Chloride	BAF	Saier et al. (unpublished results); Jentsch et al. (1995)
II. Secondary active transporters						
Major facilitator superfamily	MFS	2.1	LacY Eco	Sugars, drugs, metabolites, neurotransmitters, organic cations, nucleosides, carboxylates, organic and inorganic anions	B	Pao et al. (1998); Paulsen & Skurray (1993); Paulsen et al. (1996a,b)
Glycoside-pentose-hexuronide translocation symporter	GPH	2.2	MelB Eco	Glycosides, pentoses, hexuronides	B	Poolman et al. (1996)
Amino acid-polyamine-choline symporter	APC	2.3	PheP Eco	Amino acids, cholines, polyamines	BAE	Reizer et al. (1993)
Cation diffusion facilitator	CDF	2.4	CzrD Acu	Metal ions	BAE	Paulsen & Saier (1997)
Resistance-nodulation-cell division	RND	2.6	AcrB Eco	drugs, metal ions, lipooligosaccharides	B	Paulsen et al. (1996b)
Small multidrug resistance	SMR	2.7	Smr Sau	Drugs	B	Paulsen. et al. (1996b)
Gluconate:H+ symporter	GntP	2.8	GntP Eco	Gluconate	B	Peekhaus et al. (1997)
L-rhamnose transporter	RhaT	2.9	RhaT Eco	Rhamnose	B	Tate et al. (1992)
2-keto-3-deoxygluconate transporter	KdgT	2.10	KdgT Eco	2-Keto-3-deoxygluconate	B	Allen et al. (1989)
Citrate-Mg2+:H+ symporter	CitMHS	2.11	CitM Bsu	Citrate	C+	Boorsma et al. (1996)
C4-dicarboxylate transporter	Dcu	2.13	DcuA Eco	Dicarboxylates	G-	Saier et al. (unpublished results)
Lactate permease	LctP	2.14	LctP Eco	Lactate	BA	Saier et al. (unpublished results)
Betaine/carnitine/choline transporter	BCCT	2.15	BetT Eco	Betaine, carnitine, choline	B	Saier et al. (unpublished results)
Tellurite-resistance/dicarboxylate transporter	TDT	2.16	TehA Eco	Tellurite, dicarboxylates	BA	Saier et al. (unpublished results)
Proton-dependent oligopeptide transporter	POT	2.17	Ptr2 See	Peptides	BE	Saier et al. (unpublished results)
Ca2+:cation antiporter	CaCA	2.19	YrbG Eco	Ca2+/Na+ or H+	BAE	Saier et al. (unpublished results)
Inorganic phosphate transporter	PiT	2.20	PitA Eco	Phosphate	BAE	Saier et al. (unpublished results)
Sodium-solute symporter	SSS	2.21	PutP Eco	Amino acids, sugars, vitamins, nucleosides, inorganic anions	BE	Reizer et al. (1994)
Neurotransmitter:sodium symporter	NSS	2.22	Gat1 Hsa	Neurotransmitters	BAE	Reizer et al. (1994)
Dicarboxylate/amino acid:cation (Na+ or H+) symporter	DCS	2.23	GltP Eco	Dicarboxylates, amino acids	BE	Saier et al. (unpublished results)
Citrate:Na+ symporter	CitS	2.24	CitS Kpn	Tri- and dicarboxylates	B	Kawai et al. (1997)
Alanine:Na+ symporter	ASS	2.25	DagA Alu	Alanine	B	Reizer et al. (1994)
Branched chain amino acid:Na+ symporter	LIVSS	2.26	BrnQ Sty	Branched chain amino acids	B	Reizer et al. (1994)
Glutamate:Na+ symporter	ESS	2.27	GltS Eco	Glutamate	G-	Reizer et al. (1994)
Bile acid:Na+ symporter	BASS	2.28	NhaA Rno	Bile acids	BE	Saier et al. (unpublished results)
NhaA Na+:H+ antiporter	NhaA	2.33	NhaA Eco	Na+/H+	G-	Gerchman et al. (1993)
NhaB Na+:H+ antiporter	NhaB	2.34	NhaB Eco	Na+/H+	G-	Pinner et al. (1992)
NhaC Na+:H+ antiporter	NhaC	2.35	NhaC Bfi	Na+/H+	B	Ivey et al. (1991)
Monovalent cation:proton antiporter-1	CPA-1	2.36	Nhe1 Rno	Na+/H+	BAE	Saier et al. (unpublished results); Orlowski et al. (1992)
Monovalent cation:proton antiporter-2	CPA-2	2.37	KefC Eco	K+, Na+/H+	BAE	Saier et al. (unpublished results); Stumpe et al. (1996)
K+:H+ transporter	Trk	2.38	Trk1/AF Eco	K+/H+	BAF	Saier et al. (unpublished results); Stumpe et al. (1996)

Name	Abbreviation	TC no.[a]	Gene designation[b]	Substrate	Group[c]	Reference
Nucleobase:cation symporter-1	NCS1	2.39	CodB Eco	Cytosine, uracil, allantoin	BE	Saier et al. (unpublished results)
Nucleobase:cation symporter-2	NCS2	2.40	UraA Eco	Nucleosides	BE	Saier et al. (unpublished results)
Nucleoside uptake permease	NUP	2.41	NupG Eco	Nucleosides	BE	Saier et al. (unpublished results)
Aromatic amino acid permease	ArAAP	2.42	Mtr Eco	Amino acids	G−	Sarsero & Pittard (1997)
Serine/threonine permease	STP	2.43	SdaC Eco	Amino acids	G−	Shao et al. (1994)
Formate:nitrite porter	FNP	2.44	FocA Eco	Formate, nitrite	BA	Saier et al. (unpublished results)
Metal ion transporter	MIT	2.45	CorA Eco	$Mg^{2+}, Co^{2+}, Mn^{2+}, Al^{3+}$	BA	Smith et al. (1993)
Benzoate:H+ symporter	BenE	2.46	BenE Aeo	Benzoate	G−	Neidle et al. (1991)
Divalent anion:Na+ symporter	DASS	2.47	NaDC1 Hsa	Sulfate/phosphate/carboxylates	BAE	Markovich et al. (1993)
Ammonium transporter	Amt	2.49	AmtB Eco	Ammonium	BAE	Saier et al. (unpublished results)
Ni2+:Co2+ transporter	NiCoT	2.52	HoxN Reu	Ni^{2+}, Co^{2+}	B	Wolfram et al. (1995)
Sulfate transporter	SulP	2.53	Sul1 Sce	Sulfate	BE	Saier et al. (1995); Smith et al. (unpublished results)
Tripartite ATP-independent periplasmic transporter	TRAP-T	2.56	DctPQM Rca	Dicarboxylates	B	Forward et al. (1997)
Phosphate:Na+ symporter	PNaS	2.58	NapT Rno	Phosphate	BE	Magagnin et al. (1993)
III. Primary active transporters						
ATP-binding cassette	ABC	3.1	MalEFGK Eco	Sugars, amino acids, drugs, metal ions, Metabolites, peptides, vitamins, proteins, complex carbohydrates	BAE	Kuan et al. (1995)
H+- or Na+-translocating F-type, V-type and A-type ATPase	F-ATPase	3.2	UncA-H Eco	H^+/Na^+	BAE	Blair (1996)
Cation-translocating P-type ATPase	P-ATPase	3.3	CadA Sau	M^+, M^{2+}	BAE	Fagan & Saier (1994)
Arsenical efflux	Ars	3.4	ArsAB Eco	Arsenite, antimonite	BAE	Tsai (1997)
IV. Group translocators						
PTS glucose-glucoside	Glc	4.1	PtsG/Crr Eco	Glucose, sucrose, β-glucosides, N-acetylglucosamine	B	Reizer et al. (1996c)
PTS fructose-mannitol	Fru	4.2	FruAB Eco	Fructose, mannitol	B	Reizer et al. (1996c)
PTS lactose-cellobiose	Lac	4.3	CelABC Eco	Lactose, cellobiose	B	Reizer et al. (1996a)
PTS glucitol	Gut	4.4	GutAB Eco	Glucitol	B	Reizer et al. (1996b)
PTS galactitol	Gat	4.5	GatABC Eco	Galactitol	G−	Nobelmann & Lengeler (1996)
PTS mannose-fructose-sorbose	Man	4.6	ManXYZ Eco	Mannose, sorbose, fructose	B	Reizer et al. (1996d)
V. Unclassified						
Polysaccharide transporter	PST	99.1	RfbX Eco	Polysaccharides	B	Paulsen et al. (1997)
MerTP mercuric ion (Hg2+) permease	MerTP	99.2	MerTP Sau	Mercury and heavy metal ions	B	Silver & Phung (1996)
Nicotinamide mononucleotide uptake permease	NMN	99.4	PnuC Eco	Nicotinamide mononucleotide	G−	Foster et al. (1990)
K+ uptake permease	Kup	99.5	Kup Eco	Potassium	BE	Bañuelos et al. (1995)
L-lysine exporter	LysE	99.6	LysE Cgl	Lysine	B	Vrljic et al. (1996)
Chromate ion transporter	CIT	99.7	ChrA Aeu	Chromium ions	BA	Nies & Silver (1995)
Ferrous iron uptake	FeoB	99.8	FeoB Eco	Ferrous iron	BA	Kammler et al. (1993)
Mg2+ transporter	Mgt	99.19	MgtE Fst	Mg^{2+}, Co^{2+}	BA	Townsend et al. (1995)

[a] TC no., classification number of the transport commission (Saier, 1998).

[b] The cited example indicates the abbreviation of the protein, based on the encoding gene designation (first four letters), followed by a three letter abbreviation of the source organism (e.g. Eco, E. coli; see text for detail).

[c] The occurrence of members of the family within the following phylogenetic groups of organisms is indicated as follows: Gram-negative bacteria, G −; Gram-positive bacteria, G +; bacteria, B; Archaea, A; Eukaryotes, E.

channel of *E. coli* (Häse *et al.*, 1995; Sukharev *et al.*, 1994, 1996) and protein translocating holins (Young & Bläsi, 1995). The remaining polytopic membrane proteins, possessing four or more putative TMSs, include recognizable cytoplasmic membrane transport proteins (about half) as well as integral membrane receptors, electron carriers, and enzymes.

Our analyses (see below) indicate that most established solute transport proteins possess six or more TMSs. As these proteins possess the lowest percentage of unique proteins or proteins of unknown function (unpublished observations), we conclude that the majority of transport proteins have functionally identified homologs.

M. jannaschii was found to possess a larger percentage of orphan proteins for almost all topological types compared with the other organisms examined (data not shown). This fact presumably reflects the meager amount of sequence and functional data available for the archaea as compared with bacteria and suggests that archaeal-specific families will be identified as additional sequence and functional data become available.

Identification and classification of families of cytoplasmic membrane transport proteins

Sequence comparison searches of the chromosomally encoded putative integral membrane proteins from each of the organisms examined enabled us to identify all homologs of functionally characterized and sequenced transport proteins. These proteins were classified on the basis of sequence similarities into 67 distinct families. The complete catalog of transporters according to organism and family (including function and functional prediction) as well as the catalog of all identified transporter families are publically available on our World Wide Web sites (see Materials and Methods). It is our intention to maintain and update these sites in the future as additional genomic sequence data become available and as new functional information is published. Individuals wishing to provide corrections or additional information, particularly regarding novel functionally identified transporters, are encouraged to contact us *via* the e-mail addresses available through our WWW sites.

Much of the information currently available through our web sites is summarized in Tables 2 and 3. Table 2 presents the various transporter families, grouped according to transport mode and energy coupling mechanism. For each family, a family designation, its abbreviation, its transport commission (TC) catalog number, and a representative, well characterized member are provided. The Table also includes a compilation of typical substrates transported by the protein members of the family and an indication of the kingdoms (bacterial (B), eukaryotic (E) and archaeal (A)) in which members of the family have been found. A reference is cited which provides a phylogenetic

description of the family (when available) or provides a description of either the family or a well characterized member of the family.

Bacterial channel proteins

Channel proteins (class I families) include members of the large MIP family of aquaporins and glycerol facilitators (Park & Saier, 1996) as well as members of several ion channel protein families. Members of both voltage-sensitive and mechanosensitive ion channel families are present in the bacteria under study, and both cation-selective and anion-selective channels are represented (Jentsch *et al.*, 1995; Konings *et al.*, 1996). Except for the small prokaryotic-specific mechanosensitive ion channels (Sukharev *et al.*, 1994, 1996; Häse *et al.*, 1995), these families are represented in eukaryotes as well as bacteria. While all of the bacteria examined possess ion channel proteins, members of no channel protein family are found in all of these organisms.

Secondary carriers

Facilitators and secondary active transporters (class II families) represent the largest and most diverse category of transporters. Of the 67 families included in Tables 2 and 3, 45 fall into this category. The largest of these families are the ubiquitous major facilitator superfamily (MFS; Henderson & Maiden, 1990; Griffith *et al.*, 1992; Marger & Saier, 1993; Essenberg *et al.*, 1997; Goswitz & Brooker, 1993, 1995; Pao *et al.*, 1998) and the ubiquitous amino acid-polyamine choline (APC) family (Reizer *et al.*, 1993). While hundreds of MFS and dozens of APC transporters have been sequenced, relatively few members of any one of the other secondary transporter families represented have been sequenced. These families typically include from 1 to 20 currently sequenced members.

The secondary transporters listed in Table 2 fall into three general groups: (1) the proton motive force (pmf)-driven group of solute:proton symporters; (2) the sodium motive force (smf)-driven group of solute:Na$^+$ symporters, and (3) the so-called metal ion exchangers. Families with a preponderance of pmf-dependent permease members may also include members that catalyze sodium symport, uniport and/or solute:solute antiport. Similarly, select members of a family of smf-dependent permeases may catalyze pmf-dependent transport. Thus, pmf and smf-dependent systems seem to have undergone facile interconversion throughout evolutionary history. Finally, the metal ion exchangers may also catalyze uniport or symport with protons in addition to metal ion:proton antiport.

Examination of the occurrence of members of the 45 secondary facilitator families represented in Table 2 reveals that of them, 21 (48%) are bacterial-specific. Of these, six (14%) are found exclusively

Table 3. Distribution of constituent members of transport protein families according to organism

Protein family	Number of transport systems						
	E. coli	*H. influenzae*	*H. pylori*	*B. subtilis*	*M. genitalium*	*Synechocystis*	*M. jannaschii*
I. Channel proteins							
MIP	2	2	0	1	1	1	0
Mscl	1	1	0	1	0	1	0
VIC	1	0	0	1	0	1	2
ClC	2	0	1	0	0	2	1
Subtotal	6	3	1	3	1	5	3
II. Secondary active transporters							
MFS	64	11	7	65	1	4	2
GPH	7	0	0	3	0	1	0
APC	22	3	2	18	2	1	2
CDF	2	0	0	4	0	0	1
RND	6	1	3	1	0	4	0
SMR	5	0	0	5	0	0	0
GntP	7	2	0	2	0	0	0
RhaT	1	0	0	1	0	0	0
KdgT	1	0	0	1	0	0	0
CitMHS	0	0	0	3	0	0	0
Dcu	2	2	1	0	0	0	0
LctP	2	1	2	2	0	0	0
BCCT	3	1	0	1	0	0	0
TDT	1	1	0	0	0	0	2
POT	4	0	0	1	0	0	0
CaCA	2	0	0	1	0	2	1
PiT	3	1	1	2	0	0	1
SSS	4	3	1	5	0	1	0
NSS	0	2	2	2	0	0	1
DCS	4	1	0	4	0	0	0
CSS	0	0	0	2	0	0	0
ASS	1	2	0	4	0	0	0
LIVSS	1	1	0	2	0	0	0
ESS	1	1	1	0	0	0	0
BASS	0	0	0	1	1	0	1
NhaA	1	1	1	0	0	0	0
NhaB	1	1	0	0	0	0	1
NhaC	0	2	0	2	0	0	0
CPA-1	2	0	0	1	0	2	2
CPA-2	3	1	2	2	0	3	0
Trk	2	1	0	2	2	1	1
NCS1	2	0	0	2	0	1	0
NCS2	5	1	0	4	0	0	0
NUP	3	1	1	3	0	0	0
ArAAP	5	1	0	0	0	0	0
STP	3	1	1	0	0	0	0
FNP	3	1	0	2	0	0	0
MIT	1	1	1	2	0	2	1
BenE	1	0	0	0	0	0	0
DASS	3	2	2	1	0	0	1
Amt	1	0	0	1	0	3	2
NiCoT	0	0	1	0	0	0	0
SulP	1	0	1	2	0	4	0
TRAP-T	1	3	0	0	0	1	0
PnaS	0	0	0	1	0	0	0
Subtotal	181	50	30	155	6	30	19
III. Primary active transporters							
ABC	63	30	15	49	11	45	12
F-ATPase	1	1	1	1	1	1	1
P-ATPase	4	1	3	4	1	9	0
Ars	2	0	0	2	0	1	1
Subtotal	70	32	19	56	13	56	14
IV. Group translocators							
Glc	8	0	0	10	1	0	0
Fru	8	1	0	3	1	0	0
Lac	1	0	0	3	0	0	0
Gut	1	0	0	0	0	0	0
Gat	2	0	0	0	0	0	0
Man	2	0	0	1	0	0	0
Subtotal	22	1	0	17	2	0	0
V. Unclassified							
PST	2	0	0	2	0	0	1
MerTP	0	2	0	0	0	0	0
NMN	1	1	1	0	0	0	0
Kup	1	0	0	0	0	0	0
LysE	1	0	1	1	0	0	0
CIT	0	0	0	2	0	1	0
FeoB	1	0	1	0	0	0	0
Mgt	0	0	0	1	0	0	0
Subtotal	6	3	3	6	0	1	1
TOTAL	285	89	53	239	22	92	37

Table 4. Number of membrane transport systems compared with genome size

Organism	Genome size (Mb)	No. of transport systems/100 kb	% genes encoding transport proteins
E. coli	4.60	6.2	10.8
H. influenzae	1.83	4.8	9.8
H. pylori	1.67	3.2	5.4
B. subtilis	4.21	5.6	9.7
M. genitalium	0.58	3.7	10.2
Synechocystis PCC6803	3.57	2.6	4.7
M. jannaschii	1.66	2.2	3.5

in Gram-negative bacteria, but only one (2%) is found exclusively in Gram-positive bacteria. No family revealed by the genome analyses reported here is specific to the archaea or cyanobacteria. However, more comprehensive analyses of archaeal sequences have revealed two uniquely archaeal transporter families (unpublished results; see our second WWW site).

All but seven of the 45 secondary transporter families found encoded within the genomes of the seven prokaryotes analyzed are found in E. coli, and two of the seven families lacking from E. coli are represented in H. influenzae (Table 3). Thus, most of the secondary transporter families found in prokaryotes are represented by systems encoded within the E. coli genome and that of its close relative, H. influenzae. Four families (9%) are found in both bacteria and archaea but not in eukaryotes, while seven families (16%) are represented in both bacteria and eukaryotes but not archaea. Finally, 12 of these secondary carrier families (27%) are found in all three of the kingdoms of life. Overall, we see that 24 of the families of secondary facilitators represented in Table 2 (55%) have not been found in eukaryotes, while 29 of these families (66%) have not been found in archaea.

ATP-driven primary active transporters

In the seven prokaryotic organisms examined, only four families of solute transporters include members that are known to function as primary active transporters using ATP hydrolysis as a mode of energy coupling (class III families). These families include: (1) the ABC superfamily, capable of transporting an extremely diverse group of substances with inwardly directed or outwardly directed polarity; (2) the P-type ATPases which cat-

alyze proton, alkali and heavy metal ion transport with either inwardly or outwardly directed polarity (these enzymes are the only ones known to be autophosphorylated); (3) F, V, and A-types of ATPase multisubunit complexes that catalyze the reversible electrogenic transport of H^+ or Na^+ at the expense of ATP hydrolysis, and (4) ArsAB-type arsenite/antimonite exporters which function superficially as do members of the ABC superfamily. It should be noted that the ArsA/B-type transporters may also be able to function as secondary transporters in the absence of their ATP binding proteins (Dey & Rosen, 1995). All four of these families are ubiquitous, being found in bacteria, archaea and eukaryotes. Other ATP-dependent transport systems that translocate macromolecules (e.g. proteins, nucleic acids and complex carbohydrates) were not included in this study.

Group translocators

Group translocators (class IV families), which phosphorylate their substrates during transport, fall into the functional superfamily of phosphotransferases (members of the bacterial-specific phosphotransferase system (PTS); Postma et al., 1996). Sequence comparisons group these transporters into six phylogenetic classifications that correlate with substrate specificity. These six phylogenetic groups or families include: (1) the glucose and glucoside (Glc)-specific family, (2) the fructose and mannitol (Fru)-specific family, (3) the lactose and cellobiose (Lac)-specific family, (4) the glucitol (Gut)-specific family, (5) the galactitol (Gat)-specific family, and (6) the mannose-fructose-sorbose (Man)-specific family. These multicomponent transporters consist of three or four proteins or protein domains (IIA, B and C,

Table 5. Comparison of the energy coupling mechanisms in each organism

	Primary active transport (%)	Secondary active transport (%)	Group translocation (%)	Energy-independent channel transport (%)	Unclassified (%)
E. coli	25	64	8	2	2
H. influenzae	36	56	1	3	3
H. pylori	36	57	0	2	6
B. subtilis	23	65	7	1	3
M. genitalium	62	24	10	5	0
Synechocystis	61	32	0	5	1
M. jannaschii	39	44	0	8	8

and sometimes D), which clearly evolved from a variety of protein sources (Saier & Reizer, 1992, 1994). At least some of these different families apparently evolved independently of each other. Thus, the IIA constituents of the Glc and Fru systems evolved from different sources and have completely different three-dimensional structures, but the IIC constituents are homologous. By contrast, all constituents of the Man family evolved independently of those of the other five families. For these reasons, the PTS is said to consist of "mosaic" transport systems with a unified mechanism and function, but without a unified structure or evolutionary source (Saier & Reizer, 1994).

All of the PTS transporters utilize phosphoenolpyruvate (PEP) as both the energy source for transport and the phosphoryl donor for sugar phosphorylation. The general energy coupling proteins (Enzyme I and the heat stable phosphocarrier protein, HPr) pass the phosphoryl group from PEP to the sugar-specific IIA and IIB domains or proteins in preparation for group translocation catalyzed by the IIC constituent of the Enzyme II complex (Saier & Reizer, 1992, 1994). To date, no PTS homolog has been identified in an archaeon or a eukaryote (J. Reizer and M.H.S., Jr, unpublished observations). The results show that not only the archaeon, *M. jannaschii*, lacks genes homologous to PTS-encoding genes, but that the blue green bacterium, *Synechocystis*, and the Gram-negative bacterium, *H. pylori*, also lack such genes.

Transporters of unknown classification

Several families are listed among class V families (families of unknown classification), because, although their transport functions are established, their modes of transport or energy coupling mechanisms are not. Many of these families are likely to prove to fall into the class II group of secondary active transporters, but a few may prove to be members of class I (channel-type) or class III (ATP-dependent). Finally, additional polytopic proteins may be transporters, but no evidence for this fact is currently available because of a lack of functionally characterized homologs. Proteins in these families are not included in Tables 3 and 4 and can only be identified when functional data become available.

Transporter family representation in seven prokaryotes

Table 3 provides the numbers of transport system paralogs for each of the 67 families encoded within the genomes of the seven organisms examined. Interestingly, with the exception of *M. jannaschii* and *Synechocystis* PCC6803, the total number of transport systems per organism is approximately proportional to the genome size despite the tremendous disparities in the sizes of the genomes analyzed (eight-fold; see Table 4). Thus, there is less than two-fold variation among the five

eubacteria examined and less than three-fold variation when all seven organisms are considered. This fact is emphasized by the results summarized in Table 4 which presents the number of transporters in each organism per 100 kb of genomic DNA. As can be seen, the five eubacteria under study possess between 3.2 and 6.2 transporters per 100 kb. *M. jannaschii* and *Synechocystis* PCC6803 exhibit approximately twofold higher values (2.2 to 2.6). This last observation we believe reflects both the ecological niches and the metabolic capabilities of the latter bacteria which live in nutrient poor marine and fresh water environments, respectively, and derive their energy primarily from metabolism of exogenous inorganic compounds (Rippka *et al.*, 1979; Bult *et al.*, 1996; Kaneko *et al.*, 1996).

Comparing the Gram-negative *H. influenzae* (1.83 Mbp) or *H. pylori* (1.67 Mpb) with *E. coli* (4.60 Mbp), or the Gram-positive *M. genitalium* (0.58 Mbp) with *B. subtilis* (4.21 Mbp), it is clear that the increased numbers of transport systems in organisms with large genomes are due both to increased numbers of paralogs within the representative large families (e.g. ABC and MFS) and to increased representation of smaller families. The increase in small family representation is comparable to the increase in large family representation both in the Gram-positive bacterial pair and in the three Gram-negative bacteria examined.

Distribution of transporter types in the different microorganisms

Comparison of the numbers of transporters within each of the four energy coupling categories included in Table 3 reveals dramatic differences among the various organisms examined (Table 5). Thus, *E. coli* and *B. subtilis* possess comparable distributions of transporter types, with emphasis on secondary active transporters, but *H. influenzae*, *H. pylori* and *M. genitalium* show a relative increase in ATP-dependent systems, particularly ABC-type transporters. This tendency is particularly pronounced in *M. genitalium* which has the smallest genome and lacks both electron transport and a functional TCA cycle (Fraser *et al.*, 1995). This last observation accounts for the reliance of *M. genitalium* on primary active transport as this organism derives its energy exclusively from substrate level phosphorylation.

Examination of *M. jannaschii*, the only archaeon represented, reveals a pattern quite different from those observed for the bacteria. First, there is a greater reliance on ATP-dependent systems, particularly ABC-type systems, compared to *E. coli* or *B. subtilis* (Tables 3 and 5). Second, there are no PTS group translocators. The PTS at present appears to be eubacterial specific as no such sequences have been revealed during extensive sequencing of archaeal and eukaryotic genomes (J. Reizer and M.H.S., Jr, unpublished observations). Third, putative cation-selective and anion-selective channels are both present in *M. jannaschii*, a situ-

ation not observed for the other sequenced small genome organisms. Finally, *M. jannaschii* exhibits a larger percent of putative transporters of unknown function than any of the bacteria except *Synechocystis*

Synechocystis PCC6803 resembles *M. genitalium* in exhibiting a much greater proportion of ATP-dependent transporters relative to secondary transporters. Like *H. pylori* and *M. jannaschii*, it lacks PTS group translocators. However, it possesses the full complement of channel proteins found in *E. coli*. These channels include a MIP family protein that may function in Cu^{2+} homeostasis (Kashiwagi *et al.*, 1995; Park & Saier, 1996), a putative mechanosensitive ion channel protein, and putative cation and anion-selective ion channels (Table 3).

H. influenzae shows some interesting tendencies relative to its close relative, *E. coli*. First, it exhibits an increased proportion of ABC-type permeases relative to secondary transporters as noted above. Second, it shows a dramatic decrease in the proportion of pmf-dependent secondary transporters but a large increase in the proportion of metal ion exchangers, and a particularly large increase in the proportion of sodium-dependent nutrient uptake systems. Finally, there is a marked decrease in the proportion of PTS-type systems. We expect that these observations reflect multiple ecological and metabolic considerations. For example, *H. influenzae* may exist primarily in a high salt, carbohydrate poor medium where many nutrients are present in low concentrations (Macfadyen & Redfield, 1996). High salt environments permit the use of Na^+ symport which can be more efficient than H^+ symport due to the greater inherent permeability of biological membranes to protons than to sodium ions. Carbohydrate deficiency would minimize the benefit of possessing an extensive PTS, and low nutrient concentrations would render high affinity ABC-type transport systems of benefit. Additionally, *H. influenzae* lacks a complete Krebs cycle, and like many other Gram-negative and Gram-positive bacteria, these organisms use glycolysis only for fructose catabolism as phosphofructokinase is lacking. This latter fact explains why only fructose is transported *via* the PTS.

The distribution of transporter types in *H. pylori* closely resembles that for *H. influenzae* except that the PTS is lacking. This last fact indicates that carbohydrate catabolism is of little importance in this organism, a suggestion substantiated by the data summarized in Table 7. *H. pylori* appears to have only one carbohydrate-specific transporter

and may depend on amino acids and carboxylates as sources of carbon.

Relative distribution of ABC- and MF superfamilies in seven prokaryotes

Table 6 summarizes the distributions of ABC-type and MFS-type transporters encoded within the genomes of the seven prokaryotes analyzed. As is apparent from the data presented, there is a tremendous difference in the relative proportions of the members of these two superfamilies, depending on organism. Thus, at one end of the scale, *B. subtilis* and *E. coli* encode within their genomes more MFS porters than ABC permease systems, while *M. genitalium* and *Synechocystis* encode more than tenfold more ABC transport systems than MFS permeases (Table 6). The other bacteria analyzed exhibit intermediate proportions of these two transporter types. The distributions of ABC and MFS transporters in the various organisms qualitatively (but not quantitatively) correlate with their relative reliances on ATP and pmf-dependent transporters, respectively. Surprisingly, the fraction of all encoded transporters that are members of these two superfamilies is almost invariant (0.38 to 0.53 for the seven prokaryotes examined; Table 6). This unexpected observation has yet to be explained.

Comparison of substrate specificities for cytoplasmic membrane transporters

Table 7 summarizes the distribution of transporters in the seven organisms analyzed according to class of compound transported. The ranges of transporter substrate specificities among the five eubacteria examined are strikingly similar. All of these bacteria can take up a wide range of organic substances as well as inorganic cations and anions. Drugs appear to be actively effluxed from them all. Interestingly, the percent of transporters that probably catalyze drug efflux is highest for the non-pathogenic soil bacterium, *B. subtilis* (17.9%), although the pathogen, *H. pylori*, comes in a close second (15.1%). The high percentage of drug transporters encoded within the genome of *B. subtilis* may render it effective in competing for growth and survival in the presence of other antibiotic producing soil microorganisms.

M. jannaschii and *Synechocystis* are fundamentally different from the eubacteria with respect to the substrate specificities of their transporters. A far greater percentage of their transporters recognize

Table 6. Distribution of ABC- *versus* MFS-type permeases encoded within the genomes of seven prokaryotes

	E. coli	H. influenzae	H. pylori	B. subtilis	M. genitalium	Synechocystis	M. jannaschii
ABC	63	30	15	49	11	45	12
MFS	64	11	7	65	1	4	2
ABC/MFS	1.0	2.7	2.1	0.7	11	11	6
Total transporters	286	89	53	239	22	92	37
ABC + MFS/Total	0.45	0.46	0.42	0.48	0.55	0.53	0.38

Table 7. Substrate specificities of transporters encoded within the genomes of seven prokaryotes

	Amino acids	Poly-amines	Peptides	Nucleosides/ pyrimidines/ purines	Nitrogenous compounds (total)	Mono/di/tri-carboxylates	Carbo-hydrates	Carbon compounds (total)	Metals	Inorganic anions	Ions (metals plus anions) (total)	Drug efflux	Other	Unknown
E. coli	14.0%	1.8%	3.9%	4.2%	**24%**	4.2%	23.9%	**28%**	9.8%	6.0%	**16%**	12.6%	2.8%	16.9%
H. influenzae	15.1%	1.2%	3.5%	3.5%	**24%**	9.3%	14.0%	**23%**	23.3%	4.7%	**28%**	5.8%	2.4%	17.5%
H. pylori	22.6%	0.0%	5.7%	3.8%	**32%**	11.3%	1.9%	**13%**	24.5%	5.7%	**30%**	15.1%	1.9%	7.6%
B. subtilis	19.8%	0.0%	1.7%	3.8%	**25%**	4.6%	15.2%	**20%**	13.3%	7.2%	**21%**	17.7%	3.3%	13.1%
M. genitalium	9.5%	4.8%	4.8%	0.0%	**19%**	0.0%	23.8%	**24%**	14.3%	4.8%	**19%**	9.5%	9.5%	19.1%
Synechocystis	7.4%	1.1%	2.1%	0.0%	**11%**	1.1%	5.3%	**6%**	31.9%	14.9%	**47%**	9.6%	11.7%	14.9%
M. jannaschii	16.7%	0.0%	0.0%	0.0%	**17%**	2.8%	0.0%	**3%**	33.3%	13.9%	**47%**	5.6%	19.4%	5.6%

inorganic ions (47% relative to 23% for the eubacteria). This fact was also noted by Clayton *et al.* (1997). Only a small number of transporters in *M. jannaschii* and *Synechocystis* appear to be involved in carbon source uptake (3 to 6% relative to an average of about 22% for the eubacteria). These differences presumably reflect the ecological niches and metabolic capabilities of *M. jannaschii* and *Synechocystis* relative to those of the other organisms analyzed. It should be noted that *M. jannaschii* and *Synechocystis* possess greater proportions of functionally unidentified transporter homologs, suggesting the existence of novel types of transport systems that are as yet uncharacterized.

Transporters of *E. coli* and *B. subtilis* exhibit similar ranges of specificity. *H. influenzae* is similar in its transport capability to *E. coli* but exhibits an increased proportion of organic anion porters with a corresponding decrease in carbohydrate porters. Metal ion homeostasis may be of particular importance in this organism as an increased proportion of metal ion transporters (Table 7) correlates with the high proportion of solute:Na^+ (*versus* H^+) symporters (see Table 3).

M. genitalium possesses only a small number of transporters, reflective of its small genome size, and it exhibits a higher percentage of unknown transporters than the other eubacteria examined (Table 7). Conclusions based on comparisons with this organism should be treated with caution. *M. genitalium* has an abbreviated array of membrane transporters with most potentially non-essential permease systems eliminated. The absence of carboxylate transporters correlates with its lack of a Krebs cycle and an electron transfer chain. However, the absence of recognizable transporters for precursors of nucleic acids suggests that one or more of the unrecognized transport systems must transport these substances.

Discussion

Evolutionary considerations

This paper represents the first comprehensive attempt to identify and classify all recognizable transporters encoded within the sequenced genomes of prokaryotic organisms. Analyses of seven genomes have enabled us to identify 67 distinct families of transport systems based on primary structure. These families include large, well-characterized families such as the major facilitator (MF) and ATP-binding cassette (ABC) superfamilies, but they also include many smaller, hitherto unrecognized families. These small families exhibit a large spectrum of properties and organismal distributions. For example, while some (about half) are restricted to bacteria and exhibit extensive sequence conservation, others are widely distributed in prokaryotic and eukaryotic phyla (42%) and exhibit tremendous sequence diversion. On this basis, we estimate that some of the latter families are probably ancient families that date back over

three billion years. Nevertheless, some of these ancient families are very restricted with respect to their substrate specificities. Thus, proteins of the CDF family (TC no. 2.4) transport only divalent cations; those of the CaCA family (TC no. 2.19) are apparently specific for Ca^{2+}; those of the Pit family (TC no. 2.20) are specific for inorganic phosphate, and those of the Amt family (TC no. 2.49) transport only NH_4^+. It is noteworthy that most of the ubiquitous families with narrow substrate specificities transport inorganic ions. However, some families which include members specific for inorganic ions such as the MF, RND and ABC families (TC nos 2.1, 2.6 and 3.1, respectively) transport many other substrates in addition to inorganic ions. Other families are specific for restricted types of organic substrates (e.g. the RhaT, KdgT and LctP families; TC nos 2.9, 2.10 and 2.14, respectively), but these small families have been found only in a limited range of organisms (see Table 2). Whether these families have recently appeared or have ancient roots is an interesting question that may become answerable when more sequence data become available.

Architecture and evolutionary potential of transport proteins

The two largest transporter families found in nature, the ABC superfamily and the MFS, are particularly interesting with respect to their remarkable substrate ranges. Thus, proteins of the MFS are capable of transporting almost all types of low molecular weight compounds (e.g., sugars, drugs, inorganic and organic anions and cations, etc.). However, the ABC superfamily can transport all of these substrates as well as macromolecules such as proteins, complex carbohydrates and lipids. To date, no member of the MFS has been found to transport a molecule of >1000 Da size. We suspect that this difference in substrate range may be attributable to the dimensions and flexibility of the transmembrane permease channels. Thus, the channels of ABC permeases may be more flexible and accommodating than those of the MFS permeases. The constituent α-helices may be able to "breathe," and the subunits may even dissociate from one another. We expect that these two permease types will prove to exhibit very different channel architectures.

The broad specificities and variable polarities observed for members of the ABC and MF superfamilies presumably account for the large numbers of paralogs found within a single bacterium. *E. coli*, for example, has 64 MFS and 63 ABC transport systems, and these *E. coli* paralogs exhibit nearly the entire gamut of substrate specificities represented within these two superfamilies in all living organisms (see our web sites for descriptions of the individual permease systems).

Some smaller families, such as the APC (TC no. 2.3) and SSS (TC no. 2.21) families include members that exhibit an intermediate range of substrate

specificities. For example, APC family members transport amino acids, polyamines and choline by proton symport and/or by solute:solute antiport, while the SSS proteins transport amino acids, sugars, vitamins, nucleosides and inorganic anions, all by Na^+ symport. The variability in the substrate specificities of APC family permeases is easy to rationalize as all of the substrates are structurally related. However, the same is not true of the SSS permeases. In this latter case, the only common feature appears to be recognition and cotransport of Na^+. The variability of the substrates recognized by the different SSS family members (e.g. proline, glucose, pantothenate, adenosine, iodide, etc.) is difficult to explain. Although this family is widespread and diverse in substrate specificity, it remains a relatively small family with few currently sequenced members and much less sequence diversion than, for example, the MFS. This limited distribution correlates with the uniformity exhibited by its members with respect to polarity of transport and the nature of the cotransported cation. We anticipate that SSS family proteins will prove to exhibit a channel architecture that is different from that of either MFS or ABC-type transport systems.

Phylogeny and substrate specificity

As noted above, the SSS appears to be an exceptional family. A primary theme revealed by our analyses of permease families is the striking correlation of phylogeny with substrate specificity. This observation clearly suggests that substrate specificity has been a well conserved trait throughout evolutionary history, almost irrespective of the evolving family. At first glance this fact seems to be at odds with recent experimental work showing that single amino acid substitutions can dramatically alter the substrate specificity and even the polarity of transport. For example, a single amino acyl substitution in bacteriorhodopsin of *Halobacterium salinarium* (Asp 85 → Thr) changes the protein from a proton pump to a chloride pump of opposite polarity (Sasaki *et al.*, 1995). Moreover, single amino acid substitutions in the lactose permease of *E. coli* can change its specificity for β-galactosides so that it can transport arabinose, α-glucosides or β-glucosides, depending on the substitution (Brooker & Wilson, 1985; Collins *et al.*, 1989; King & Wilson, 1990; Olsen *et al.*, 1993; see Varela & Wilson, 1996 for a current review). One final example is the QacA drug resistance pump of *Staphylococcus aureus* which dramatically changes its range of substrates when specific point mutations are introduced (Paulsen *et al.*, 1996a,b).

On the basis of these observations, it would seem that changing substrate specificity is a simple matter. However, in none of these examples do the mutations change the class of compound recognized. It would be most interesting to conduct comparable studies with members of families that exhibit invariant specificities (e.g. the Amt family

transporters, all of which transport NH_4^+, or the CaCA family, all of which transport Ca^{2+}) to see if these permeases can also change their specificities upon introduction of point mutations. Parallel studies with the SSS family, which exhibits unusually broad substrate specificity but narrow cation specificity would be of further interest.

Energy availability, environmental factors and transport mode

We have found that some organisms exhibit a preponderance of pmf-driven transporters while others exhibit a preponderance of ATP-driven permeases. In some cases, this difference seems to be explained by the type of energy source most readily generated by the organism. *M. genitalium* generates ATP as its primary energy source and has mostly ATP-dependent permeases, while *B. subtilis* generates a pmf as its primary source of energy and exhibits the opposite tendency (Table 3). Other organisms are less readily rationalized. Thus, *E. coli* and *Synechocystis* can generate both ATP and a pmf as primary energy sources, but while *E. coli* exhibits a 2.5-fold excess of pmf-driven permeases over ATP-driven permeases, *Synechocystis* exhibits a 2.5-fold excess of ATP-driven systems. In these cases, the environments in which these organisms flourish may be important. Thus, while *E. coli* is an intestinal bacterium where nutrients are often present at high concentrations, *Synechocystis* is a freshwater organism living in nutrient-poor surroundings (Rippka *et al.*, 1979). The latter organism may require the presence of high affinity transporters to survive and grow, while low affinity uptake systems may suffice for *E. coli*. ATP-dependent nutrient transporters in general transport their solutes with much higher affinities than do pmf-driven systems. It should also be noted, however, that *Synechocystis* possesses an intracellular thylakoid membrane system that is not continuous with the plasma membrane. The generation of a pmf across the thylakoid membrane may not allow pmf-dependent uptake of nutrients across the cytoplasmic membrane.

Ubiquity of transport protein families

Only four transport protein families are represented in all seven of the organisms we have examined. These families are the MF (TC no. 2.1), APC (TC no. 2.3), ABC (TC no. 3.1) and F-type ATPase (TC no. 3.2) families. F-type ATPases probably have the same function in all living organisms: they interconvert chemical and chemiosmotic energy forms. However, examination of the protein members of the other three ubiquitous families shows that their specificities differ dramatically, and that no single solute-specific permease is present in all seven organisms. This fact stresses the transport variability observed for different microorganisms. Characterization of an organism's complement of transport proteins should reveal

many important aspects of its lifestyle: its preferred exogenous sources of energy and nutrition, its metabolic pathways, and its end products of metabolism. The realization of this goal will require extensive physiological, molecular genetic and biochemical analyses of permease function.

Transport capacity and cellular metabolic capability

M. genitalium is an organism with almost no biosynthetic potential. It must acquire all precursors for the biosynthesis of proteins, nucleic acids, lipids and complex carbohydrates from exogenous sources. One might expect that a greatly increased proportion of the genes of such an organism would encode transport proteins. In the case of *M. genitalium*, this postulate proved to be untrue. The same percent of the genomes of *M. genitalium* and *E. coli* encode recognized permeases (10.2% and 10.8%, respectively; Table 4) although *M. genitalium* has only 7% as many genes as does *E. coli*. This observation clearly suggests that many of the transport systems of *M. genitalium* must differ in specificity from those of *E. coli*, being capable of transporting many related, essential compounds. We estimate that *M. genitalium* has a single proton-translocating F-type ATPase (TC no. 3.2), a single Ca^{2+}-translocating P-type ATPase (TC no. 3.3), a single $K^+:H^+$ antiporter of the Trk family (TC no. 2.38) and two inorganic anion transporters of the ABC family (TC no. 3.1), one specific for phosphate, the other of unknown specificity. These five permeases presumably suffice to maintain the pmf and proper cytoplasmic ionic balance.

M. genitalium devotes a major fraction of its transport capacity to the task of sugar uptake for the purpose of carbon and energy acquisition. Thus, it possesses: (1) both glucose and fructose phosphotransferase systems (TC no. 4.1 and TC no. 4.2, respectively; Reizer *et al.*, 1996c); (2) two sugar permeases of the ABC superfamily, one of which probably transports oligosaccharides (TC no. 3.1.1) while the other transports monosaccharides (TC no. 3.1.2; Tam & Saier, 1993), and (3) a non-specific channel protein of the MIP family (TC no. 1.1), probably capable of transporting small, neutral, straight chain molecules such as urea, glycerol and various polyols (Park & Saier, 1996). As *M. genitalium* lacks a TCA cycle and electron flow, it is possible that it lacks specific transporters for organic carboxylates. However, even these compounds may be transported by non-specific permeases.

M. genitalium takes up amino acids *via* two APC transporters (TC no. 2.3), and it probably accumulates peptides and amines *via* two distinct ABC-type systems. These four permeases presumably provide the requisite precursors for protein synthesis. However, we did not identify permeases likely to accumulate vitamins or the precursors of lipid and nucleic acid biosynthesis. At least three transport systems of unknown specificity (two

ABC-types and one MFS-type) were identified, and these, or broad specificity sugar, amino acid or peptide permeases, may provide these functions.

Finally, *M. genitalium* possesses three transport systems that are likely to function in the active extrusion of drugs and other toxic substances (Table 7). These transport systems may serve protective roles and also allow efflux of metabolic end products. Examination of our web sites allows the reader to dissect the transport capabilities of the other bacteria analyzed. The permease complement of an organism presumably reflects its metabolic capabilities as well as the environmental conditions under which the organism evolved.

Protein topology and functional assignment

Our topological analyses suggested that in each organism studied, proteins of one to three transmembrane α-helical spanners (TMSs) possess fewer identifiable homologs than proteins of 0 or ≥4 TMSs (unpublished results). This observation presumably relates to the functions of these proteins. We suggest that a large percentage of these proteins primarily serve structural functions, and that specific amino acyl residues within these proteins are, in general, far less critical to function than are those of soluble enzymes (0 TMS) and catalytic transport proteins (usually >4 TMSs). Lack of constraint for sequence divergence can explain our unpublished observations.

Conclusions and perspectives

The results described here summarize analyses that substantially extend our understanding of the transport capabilities of microorganisms and the evolution of transport proteins. Further molecular genetic, biochemical and physiological analyses will be required to identify the functions of the permeases and their families that are as yet uncharacterized. Further genome sequencing projects should additionally reveal the distributions of the various types of transport systems in the living world. Such studies together with three-dimensional analyses will ultimately provide an understanding of the evolutionary origins and mechanistic details of transport systems of all types.

Materials and Methods

The complete protein inventories from each organism were obtained from the following sites: *H. influenzae*, *M. genitalium*, *H. pylori* and *M. jannaschii* (http://www.ti-gr.org/tdb/mdb/mdb.html; Fleischmann *et al.*, 1995; Fraser *et al.*, 1995; Bult *et al.*, 1996; Tomb *et al.*, 1997); *E. coli* (http://www.genetics.wisc.edu:80/index.html; Blattner *et al.*, 1997), and *Synechocystis* PCC6803 (http://www.kazusa.or.jp/cyano/cyano.html; Kaneko *et al.*, 1996). The *B. subtilis* sequences analyzed were obtained courtesy of A. Danchin prior to release of the complete NRSub database (Biaudet *et al.*, 1996; Kunst *et al.*, 1997; http://www.pasteur.fr/bio/SubtiList.html). The com-

plete genomic sequence of *M. pneumoniae* (Himmelreich *et al.*, 1996) revealed a complement of transporters similar to that of *M. genitalium*, and analysis is consequently not presented here. Analyses of the *S. cerevisiae* genome (Horák, 1997) will be reported elsewhere (unpublished).

Automated hydropathy analyses on the complete complement of proteins encoded within each genome were performed using a modified version of the program MEMSAT (Jones *et al.*, 1994, modified by us). Due to limitations of the program, proteins which were less than 50 amino acids in length were excluded from the analyses, and proteins in excess of 1000 amino acids were truncated. Additional hydropathy analyses were performed using the programs TOP-PRED (Claros & von Heijne, 1994) and TMpred (Hofmann & Stoffel, 1993). Based on these analyses, the number of TMSs in each protein analyzed was predicted.

All of the proteins, which contain ⩾1 predicted TMS encoded within each genome, were screened against the databases for sequence similarities with known and putative transport proteins. Database searches were initially performed using BLAST (Altschul *et al.*, 1990). When no significant sequence similarity was detected, additional searches were performed using FASTA (Lipman & Pearson, 1985) and BLITZ (Sturrock & Collins, 1993). The significance of sequence similarities detected was confirmed using the programs RDF2 (Pearson & Lipman, 1988) and GAP (Devereux *et al.*, 1984) with 200 random shuffles. A comparison score of nine standard deviations obtained with either one of these programs was considered sufficient to establish homology (Doolittle, 1986; Saier, 1994). Based on these analyses, the potential membrane transporters identified were classified according to sequence similarities into separate families of transporters.

Identification of transporters from each organism was cross checked against the protein assignments from the genome sequencing projects and from earlier relevant studies (e.g. Riley, 1993; Riley & Lebelan, 1996). Additionally, representative members of each prokaryotic family of transporters and of various eukaryotic-specific transporter families were used to rescreen the genomes of each organism translated in all six reading frames. Auxiliary proteins and hydrophilic components of particular transport systems were subsequently identified by database searches. Functional predictions for each putative transport protein were made whenever possible based on sequence and phylogenetic analyses of the families of transporters. Phylogenetic trees were constructed using the TREE (Feng & Doolittle, 1990) and PILEUP (Devereux *et al.*, 1984) programs. For each organism, the family classifications and functional identifications/predictions for the encoded transport proteins are available at our WWW sites (http://www-biology.ucsd.edu/ ~ ipaulsen/transport/titlepage.htm); (http://www-biology.ucsd.edu/ ~ msaier/trans-port/titlepage.html).

It should be noted that the analyses reported here do not include: (1) outer membrane transport proteins (Jeanteur *et al.*, 1991); (2) proteins such as those of the *E. coli* TonB/ExbB/ExbD complex which transduce energy from the inner to the outer membrane to drive outer membrane transport processes (Braun *et al.*, 1994); (3) proteins involved in most protein secretory pathways, e.g. the *E. coli* Sec proteins (Saier *et al.*, 1989); (4) proton and sodium ion-translocating electron transfer proteins (Gennis & Stewart, 1996); (5) Na⁺-transporting carboxylic acid decarboxylases (Dimroth, 1997); (6) ion-

translocating flagellar motor (Mot) proteins (Nguyen & Saier, 1996); and (7) proteins involved in competence for DNA uptake (Dubnau, 1991; Macfadyen *et al.*, 1996). Many of these proteins are, however, included in the global classification of transport proteins included in the second of our two web sites cited above and to be published elsewhere (Saier, 1998). Auxiliary transport proteins, such as proteins from the MFP family (Dinh *et al.*, 1994) or the membrane-periplasmic auxiliary proteins of the MPA1 and MPA2 families (Paulsen *et al.*, 1997) were not considered as separate transport systems, but were treated as components of the transporters with which they function.

References

Alexander, S. P. H. & Peters, J. A. (1997). *1997 Receptor and Ion Channel Nomenclature Supplement* (Babbedge, R., ed.), eighth edit., pp. 76–84, Elsevier Trends Journals, Cambridge, UK.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Allen, C., Reverchon, S. & Robert-Baudouy, J. (1989). Nucleotide sequence of the *Erwinia chrysanthemi* gene encoding 2-keto-3-deoxygluconate permease. *Gene*, **83**, 233–241.

Bañuelos, M. A., Klein, R. D., Alexander-Bowman, S. J. & Rodríguez-Navarro, A. (1995). A potassium transporter of the yeast *Schwanniomyces occidentalis* homologous to the Kup system of *Escherichia coli* has a high concentrative capacity. *EMBO J.* **14**, 3021–3027.

Biaudet, V., Samson, F., Anagnostopoulos, C., Ehrlich, S. D. & Bessières, P. (1996). Computerized genetic map of *Bacillus subtilis. Microbiology,* **142**, 2669–2729.

Blair, A., Ngo, L., Park, J., Paulsen, I. T. & Saier, M. H., Jr. (1996). Phylogenetic analyses of the homologous transmembrane channel-forming proteins of the F₀F₁-ATPases of bacteria, chloroplasts and mitochondria. *Microbiology,* **142**, 17–32.

Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *277*, 1453–1474.

Boorsma, A., van der Rest, M. E., Lolkema, J. S. & Konings, W. N. (1996). Secondary transporters for citrate and the Mg²⁺-citrate complex in *Bacillus sub-*

tilis are homologous proteins. *J. Bacteriol.* **178**, 6216–6222.

Braun, V., Pilsl, H. & Groß, P. (1994). Colicins: Structures, modes of action, transfer through membranes, and evolution. *Arch. Microbiol.* **161**, 199–206.

Brooker, R. J. & Wilson, T. H. (1985). Isolation and nucleotide sequencing of lactose carrier mutants that transport maltose. *Proc. Natl Acad. Sci. USA*, **82**, 3959–3963.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J.-F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H.-P., Fraser, C. M., Smith, H. O., Woese, C. R. & Venter, J. C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058–1073.

Clark, J. A. & Amara, S. G. (1993). Amino acid neurotransmitter transporters: structure, function, and molecular diversity. *BioEssays*, **15**, 323–332.

Claros, M. G. & von Heijne, G. (1994). TopPred II: An improved software for membrane protein structure predictions. *Comput. Applic. Biosci.* **10**, 685–686.

Clayton, R. A., White, O., Ketchum, K. A. & Venter, J. C. (1997). The first genome from the third domain of life. *Nature*, **387**, 459–462.

Collins, J. C., Permuth, S. F. & Brooker, R. J. (1989). Isolation and characterization of lactose permease mutants with an enhanced recognition of maltose and diminished recognition of cellobiose. *J. Biol. Chem.* **264**, 14698–14703.

Dean, D. & Allikmets, R. (1995). Evolution of ATP-binding cassette transporter genes. *Curr. Opin. Genet. Dev.* **5**, 779–785.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.

Dey, S. & Rosen, B. P. (1995). Dual mode of energy coupling by the oxyanion-translocating ArsB protein. *J. Bacteriol.* **177**, 385–389.

Dimroth, P. (1997). Primary sodium ion translocating enzymes. *Biochim. Biophys. Acta*, **1318**, 11–51.

Dinh, T., Paulsen, I. T. & Saier, M. H., Jr (1994). A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of Gram-negaitve bacteria. *J. Bacteriol.* **176**, 3825–3831.

Doolittle, R. F. (1986). *Of urfs and orfs: A Primer on How to Analyze Derived Amino Acid Sequences*, University Science Books, Mill Valley, CA.

Dubnau, D. (1991). Genetic competence in *Bacillus subtilis*. *Microbiol. Rev.* **55**, 395–424.

Essenberg, R. C., Candler, C. & Nida, S. K. (1997). *Brucella abortus* strain 2308 putative glucose and galactose transporter gene: Cloning and characterization. *Microbiology*, **143**, 1549–1555.

Fagan, M. J. & Saier, M. H., Jr (1994). P-type ATPases of eukaryotes and bacteria: sequence comparisons and construction of phylogenetic trees. *J. Mol. Evol.* **38**, 57–99.

Fath, M. J. & Kolter, R. (1993). ABC transporters: bacterial exporters. *Microbiol. Rev.* **57**, 995–1017.

Feng, D. F. & Doolittle, R. F. (1990). Progressive alignment and phylogenetic tree construction of protein sequences. *Methods Enzymol.* **183**, 375–387.

Fischer, W.-N., Kwart, M., Hummel, S. & Frommer, W. B. (1995). Substrate specificity and expression profile of amino acid transporters (AAPs) in *Arabidopsis*. *J. Biol. Chem.* **270**, 16315–16320.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O. & Venter, J. C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.

Forward, J. A., Behrendt, M. C., Wyborn, N. R., Cross, R. & Kelly, D. J. (1997). TRAP transporters: A new family of periplasmic solute transport systems encoded by the *dctPQM* genes of *Rhodobacter capsulatus* and by homologs in diverse Gram-negative bacteria. *J. Bacteriol.* **179**, 5482–5493.

Foster, J. W., Park, Y. K., Penfound, T., Fenger, T. & Spector, M. P. (1990). Regulation of NAD metabolism in *Salmonella typhimurium*: molecular sequence analysis of the bifunctional *nadR* regulator and the *nadA-pnuC* operon. *J. Bacteriol.* **172**, 4187–4196.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P.-C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchison, C. A., III & Venter, J. C. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403.

Gennis, R. B. & Stewart, V. (1996). Respiration. In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B. W., Reznikoff, S., Riley, M., Schaechter, M. & Umbarger, H. E., eds), pp. 217–261, ASM Press, Washington, DC.

Gerchman, Y., Olami, Y., Romon, A., Taglicht, D., Schuldiner, S. & Padan, E. (1993). Histidine-226 is part of the pH sensor of NhaA, a Na⁺:H⁺ antiporter in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **90**, 1212–1216.

Goffeau, A., Park, J., Paulsen, I. T., Jonniaux, J.-L., Dinh, T., Mordant, P. & Saier, M. H., Jr (1997). Multidrug-resistant transport proteins in yeast: complete inventory and phylogenetic characterization of yeast open reading frames within the major facilitator superfamily. *Yeast*, **13**, 43–54.

Goswitz, V. C. & Brooker, R. J. (1993). Isolation of lactose permease mutants which recognize arabinose. *Membr. Biochem.* **10**, 61–70.

Goswitz, V. C. & Brooker, R. J. (1995). Structural features of the uniporter/symporter/anti-porter superfamily. *Protein Sci.* **4**, 534–537.

Griffith, J. K., Baker, M. E., Rouch, D. A., Page, M. G. P., Skurray, R. A., Paulsen, I. T., Chater, K. F., Baldwin, S. A. & Henderson, P. J. F. (1992). Membrane transport proteins: implications of sequence comparisons. *Curr. Opin. Cell Biol.* **4**, 684–695.

Häse, C. C., Le Dain, A. C. & Martinac, B. (1995). Purification and functional reconstitution of the recombinant large mechanosensitive ion channel (MscI) of *Escherichia coli*. *J. Biol. Chem.* **270**, 18329–18334.

Henderson, P. J. F. & Maiden, M. C. J. (1990). Homologous sugar transport proteins in *Escherichia coli* and their relatives in both prokaryotes and eukaryotes. *Phil. Trans. Roy. Soc. London ser. B*, **326**, 391–410.

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B. C. & Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucl. Acids Res.* **24**, 4420–4449.

Hofmann, K. & Stoffel, W. (1993). TMbase: a database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, **347**, 166.

Horák, J. (1997). Yeast nutrient transporters. *Biochim. Biophys. Acta*, **1331**, 41–79.

Ivey, D. M., Guffanti, A. A., Bossewitch, J. S., Padan, E. & Krulwich, T. A. (1991). Molecular cloning and sequencing of a gene from alkaliphilic *Bacillus firmus* OF4 that functionally complements an *Escherichia coli* strain carrying a deletion in the *nhaA* Na^+/H^+ antiporter gene. *J. Biol. Chem.* **266**, 23483–23489.

Jeanteur, D., Lakey, J. H. & Pattus, F. (1991). The bacterial porin superfamily: Sequence alignment and structure prediction. *Mol. Microbiol.* **5**, 2153–2164.

Jentsch, T. J., Gunther, W., Pusch, M. & Schwappach, B. (1995). Properties of voltage-gated chloride channels of the ClC gene family. *J. Physiol.* **482**, 19S–25S.

Jones, D. T., Taylor, W. R. & Thornton, J. M. (1994). A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry*, **33**, 3038–3049.

Kaback, H. R. (1986). Active transport in *Escherichia coli*: Passage to permease. *Annu. Rev. Biophys. Biophys. Chem.* **15**, 279–319.

Kaback, H. R., Voss, J. & Wu, J. (1997). Helix packing in polytopic membrane proteins: the lactose permease of *Escherichia coli*. *Curr. Opin. Struct. Biol.* **7**, 537–542.

Kammler, M., Schön, C. & Hantke, K. (1993). Characterization of the ferrous iron uptake system of *Escherichia coli*. *J. Bacteriol.* **175**, 6212–6219.

Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, N., Hirosawa, M., Sugiura, M., Sasamoto, S., Kimura, T., Hosouchi, T., Matsuno, A., Muraki, A., Nakazaki, N., Naruo, K., Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yasuda, M. & Tabata, S. (1996). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136.

Karp, P. D. & Paley, S. (1994). Integrated access to metabolic and genomic data. *J. Comput. Biol.* **3**, 191–212.

Karp, P. D., Riley, M., Paley, S. M. & Pellegrini-Toole, A. (1996). EcoCyc: An encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res.* **24**, 32–39.

Kashiwagi, S., Kanamaru, K. & Mizuno, T. (1995). A *Synechococcus* gene encoding a putative pore-forming intrinsic membrane protein. *Biochim. Biophys. Acta*, **1237**, 189–192.

Kawai, S., Suzuki, H., Yamamoto, K. & Kumagai, H. (1997). Characterization of the L-malate permease gene (*maeP*) of *Streptococcus bovis* ATCC 15352. *J. Bacteriol.* **179**, 4056–4060.

King, S. C. & Wilson, T. H. (1990). Identification of valine 177 as a mutation altering specificity for transport of sugars by the *Escherichia coli* lactose carrier. Enhanced specificity for sucrose and maltose. *J. Biol. Chem.* **265**, 9638–9644.

Konings, W. N., Kaback, H. R. & Lolkema, J. S., eds (1996). *Transport Processes in Eukaryotic and Prokaryotic Organisms*, Elsevier Science Publishers, Amsterdam, The Netherlands.

Krämer, R. (1994). Secretion of amino acids by bacteria: physiology and mechanism. *FEMS Microbiol. Rev.* **75**, 75–94.

Kuan, G., Dassa, E., Saurin, W., Hofnung, M. & Saier, M. H., Jr (1995). Phylogenetic analyses of the ATP-binding constituents of bacterial extracytoplasmic receptor-dependent ABC-type nutrient uptake permeases. *Res. Microbiol.* **146**, 271–278.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessières, P., Bolotin, A., Borchert, S., Borriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Bruschi, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S.-K., Codani, J.-J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Düsterhöft, A., Ehrlich, S. D., Emmerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S.-Y., Glaser, P., Goffeau, A., Golightly, E. J., Grandi, G., Guiseppi, G., Guy, B. J., Haga, K., Haiech, J., Harwood, C. R., Hénaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M.-F., Itaya, M., Jones, L., Joris, B., Karamata, D., Kasahara, Y., Klaerr-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S.-M., Levine, A., Liu, H., Masuda, S., Mauël, C., Médigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S.-H., Parro, V., Pohl, T. M., Portetelle, D., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadaie, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B.-S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tamakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wedler, E., Wedler, H., Weitzenegger, T., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H.-F., Zumstein, E., Yoshikawa, H. & Danchin, A. (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature*, **390**, 249–272.

Le, T. & Saier, M. H., Jr (1996). Phylogenetic characterization of the epithelial Na^+ channel (ENaC) family. *Mol. Mem. Biol.* **13**, 149–157.

Lee, A. G. (1996). *Biomembranes. Transmembrane Receptors and Channels*, JAI Press, Inc., Greenwich, CT.

Lipman, D. J. & Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science*, 227, 1435–1441.

Macfadyen, L. P. & Redfield, R. J. (1996). Life in mucus: sugar metabolism in *Haemophilus influenzae*. *Res. Microbiol*, 147, 541–551.

Macfadyen, L. P., Dorocicz, I. R., Reizer, J., Saier, M. H., Jr. & Redfield, R. J. (1996). Regulation of competence development and sugar utilization in *Haemophilus influenzae* and sugar utilization in *Haemophilus influenzae* Rd by a phosphoenolpyruvate:fructose phospho-transferase system. *Mol. Microbiol*. 21, 941–952.

Magagnin, S., Werner, A., Markovich, D., Sorribas, V., Stange, G., Biber, J. & Murer, H. (1993). Expression cloning of human and rat renal cortex sodium-phosphorus cotransport. *Proc. Natl Acad. Sci. USA*, 90, 5979–5983.

Maloney, P. C. (1990). Microbes and membrane biology. *FEMS Microbiol. Rev.* 87, 91–102.

Maloney, P. C. (1992). The molecular and cell biology of anion transport by bacteria. *BioEssays*, 14, 757–762.

Marger, M. D. & Saier, M. H., Jr (1993). A major superfamily of transmembrane facilitators catalyzing uniport, symport and antiport. *Trends Biochem. Sci.* 18, 13–20.

Markovich, D., Forgo, J., Stange, G., Biber, J. & Murer, H. (1993). Expression cloning of rat renal Na$^+$/SO$_4$ (2–) cotransport. *Proc. Natl Acad. Sci. USA*, 90, 8073–8077.

Mitchell, P. (1967a). Translocations through natural membranes. *Advan. Enzymol.* 29, 33–87.

Mitchell, P. (1967b). Proton-translocation phosphorylation in mitochondria, chloroplasts and bacteria: Natural fuel cells and solar cells. *FASEB J.* 26, 1370–1379.

Neidle, E. L., Hartnett, C., Ornston, L. N., Bairoch, A., Rekik, M. & Harayama, S. (1991). Nucleotide sequences of the *Acinetobacter calcoaceticus benABC* genes for benzoate 1,2-dioxygenase reveal evolutionary relationships among multicomponent oxygenases. *J. Bacteriol.* 173, 5385–5395.

Nguyen, C. C. & Saier, M. H., Jr. (1996). Structural and phylogenetic analysis of the MotA and MotB families of bacterial flagellar motor proteins. *Res. Microbiol.* 147, 317–332.

Nies, D. H. & Silver, S. (1995). Ion efflux systems involved in bacterial metal resistances. *J. Ind. Microbiol.* 14, 186–199.

Nobelmann, B. & Lengeler, J. W. (1996). Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. *J. Bacteriol.* 178, 6790–6795.

Olsen, S. G., Greene, K. M. & Brooker, R. J. (1993). Lactose permease mutants which transport (malto-)-oligosaccharides. *J. Bacteriol.* 175, 6269–6275.

Orlowski, J., Kandasamy, R. A. & Shull, G. E. (1992). Molecular cloning of putative members of the Na/H exchanger gene family. cDNA cloning, deduced amino acid sequence, and mRNA tissue expression of the rat Na/H exchanger NHE-1 and two structurally related proteins. *J. Biol. Chem.* 267, 9331–9339.

Pao, S. S., Paulsen, I. T. & Saier, M. H. (1998). Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* In the press.

Park, J. H. & Saier, M. H., Jr. (1996). Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J. Membr. Biol.* 153, 171–180.

Paulsen, I. T. & Saier, M. H., Jr (1997). A novel family of ubiquitous heavy metal ion transport proteins. *J. Membr. Biol.* 156, 99–103.

Paulsen, I. T. & Skurray, R. A. (1993). Topology, structure and evolution of two families of proteins involved in antibiotic and antiseptic resistance in eukaryotes and prokaryotes–an analysis. *Gene*, 124, 1–11.

Paulsen, I. T., Brown, M. H., Littlejohn, T. G., Mitchell, B. A. & Skurray, R. A. (1996a). Multidrug resistance proteins QacA and QacB from *Staphylococcus aureus*: membrane topology and identification of residues involved in substrate specificity. *Proc. Natl Acad. Sci. USA*, 93, 3630–3635.

Paulsen, I. T., Brown, M. H. & Skurray, R. A. (1996b). Proton-dependent multidrug efflux systems. *Microbiol. Rev.* 60, 575–608.

Paulsen, I. T., Beness, A. M. & Saier, M. H., Jr (1997). Computer-based analyses of the protein constituents of transport systems catalyzing export of complex carbohydrates in bacteria. *Microbiology*, 143, 2685–2699.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, 85, 2444–2448.

Peekhaus, N., Tong, S., Reizer, J., Saier, M. H., Jr, Murray, E. & Conway, T. (1997). Characterization of a novel transporter family that includes multiple *Escherichia coli* gluconate transporters and their homologues. *FEMS Microbiol. Letters*, 147, 233–238.

Pinner, E., Padan, E. & Schuldiner, S. (1992). Cloning, sequencing and expression of the NhaB gene, encoding a Na$^+$:H$^+$ antiporter in *Escherichia coli*. *J. Biol. Chem.* 267, 11064–11068.

Poolman, B., Knol, J., ver der Does, C., Henderson, P. J. F., Liang, W.-J., Leblanc, G., Pourcher, T. & Mus-Veteau, I. (1996). Cation and sugar selectivity determinants in a novel family of transport proteins. *Mol. Microbiol.* 19, 911–922.

Pourcher, T., Bibi, E., Kaback, H. R. & Leblanc, G. (1996). Membrane topology of the melibiose permease of *Escherichia coli* studied by *melB-phoA* fusion analysis. *Biochemistry*, 35, 4161–4168.

Reizer, J., Finley, K., Kakuda, D., MacLeod, C. L., Reizer, A. & Saier, M. H., Jr (1993). Mammalian integral membrane receptors are homologous to facilitators and antiporters of yeast, fungi and eubacteria. *Protein Sci.* 2, 20–30.

Reizer, J., Reizer, A. & Saier, M. H., Jr (1994). A functional superfamily of sodium:solute symporters. *Biochim. Biophys. Acta*, 1197, 133–166.

Reizer, J., Charbit, A., Reizer, A. & Saier, M. H., Jr (1996a). Novel phosphotransferase system genes revealed by bacterial genome analysis: operons encoding homologues of sugar-specific permease domains of the phosphotransferase system and pentose catabolic enzymes. *Genome Sci. Technol.* 1, 53–75.

Reizer, J., Mitchell, W. J., Minton, N., Brehm, J., Reizer, A. & Saier, M. H., Jr (1996b). Proposed topology of the glucitol permeases of *Escherichia coli* and *Clostridium acetobulylicum*. *Curr. Microbiol.* 33, 331–333.

Reizer, J., Paulsen, I. T., Titgemeyer, F. & Saier, M. H., Jr (1996c). Novel phosphotransferase system genes revealed by bacterial genome analysis: The com-

plete complement of *pts* genes in *Mycoplasma genitalium*. *Microbial Comp. Genomics*, **1**, 151–164.

Reizer, J., Ramseier, T. M., Reizer, A. & Saier, M. H., Jr (1996d). Novel phosphotransferase genes revealed by bacterial genome analysis: A gene cluster encoding a phosphotransferase system permease and metabolic enzymes concerned with N-acetylgalactosamine metabolism. *Microbiology*, **142**, 231–250.

Riley, M. (1993). Functions of the gene products of *Escherichia coli*. *Microbiol. Rev.* **57**, 862–952.

Riley, M. & Labelan, B. (1996). *Escherichia coli* gene products: physiological functions and common ancestries. In *Escherichia coli and Salmonella. Cellular and Molecular Biology* (Neidardt, F. C., Curtiss, R., III, Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B. W., Reznikoff, S., Riley, M., Schaechter, M. & Umbarger, H. E., eds), vol. 1, 2nd edit., pp. 2118–2202, ASM Press, Washington, DC.

Rippka, R., Deruelles, J., Waterbury, J. B., Ferdman, M. & Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* **111**, 1–61.

Saier, M. H., Jr (1994). Computer-aided analyses of transport protein sequences: Gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbial. Rev.* **58**, 71–93.

Saier, M. H., Jr (1996). Phylogenetic approaches to the identification and characterization of protein families and transport superfamilies. *Microb. Comp. Genomics*, **1**, 129–150.

Saier, M. H., Jr (1998). Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. *Advan. Microbial Physiol.* In the press.

Saier, M. H., Jr & Reizer, J. (1991). Families and superfamilies of transport proteins common to prokaryotes and eukaryotes. *Curr. Opin. Struct. Biol.* **1**, 362–368.

Saier, M. H., Jr & Reizer, J. (1992). Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvatesugar phosphotransferase system. *J. Bacteriol.* **174**, 1433–1438.

Saier, M. H., Jr & Reizer, J. (1994). The bacterial phosphotransferase system: New frontiers 30 years later. *Mol. Microbiol.* **13**, 755–764.

Saier, M. H., Jr, Werner, P. K. & Muller, M. (1989). Insertion of proteins into bacterial membranes: Mechanism, characteristics, and comparisons with the eucaryotic process. *Microbiol. Rev.* **53**, 333–366.

Sarseio, J. P. & Pittard, A. J. (1995). Membrane topology analysis of *Escherichia coli* K-12 Mtr permease by alkaline phosphatase and β-galactosidase fusions. *J. Bacteriol.* **177**, 297–306.

Sasaki, J., Brown, L. S., Chon, Y.-S., Kandori, H., Maeda, A., Needleman, R. & Lanyi, J. K. (1995). Conversion of bacteriorhodopsin into a chloride ion pump. *Science*, **269**, 73–75.

Selkov, E., Basmanova, S., Gaasterland, T., Goryanin, I., Gretchkin, Y., Maltsev, N., Nenashev, V., Overbeek, R., Panyushkina, E., Pronevitch, L., Selkov, E., Jr & Yunus, I. (1996). The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. *Nucl. Acids Res.* **24**, 26–28.

Shao, Z.-Q., Lin, R. T. & Newman, E. B. (1994). Sequencing and characterization of the *sdaC* gene and identification of the *sdaCB* operon in *Escherichia coli* K-12. *Eur. J. Biochem.* **222**, 901–907.

Silver, S. & Phung, L. T. (1996). Bacterial heavy metal resistance: new surprises. *Annu. Rev. Microbiol.* **50**, 753–789.

Smith, R. L., Banks, J. L., Snavely, M. D. & Maguire, M. E. (1993). Sequence and topology of the CorA magnesium transport systems of *Salmonella typhimurium* and *Escherichia coli*. Identification of a new class of transport protein. *J. Biol. Chem.* **268**, 14071–14080.

Smith, F. W., Hawkesford, M. J., Prosser, I. M. & Clarkson, D. T. (1995). Isolation of a cDNA from *Saccharomyces cerevisiae* that encodes a high affinity sulphate transporter at the plasma membrane. *Mol. Gen. Genet.* **247**, 709–715.

Stumpe, S., Schlösser, A., Schleyer, M. & Bakker, E. P. (1996). K+ circulation across the prokaryotic cell membrane: K+-uptake systems. In *Transport Processes and Eukaryotic and Prokaryotic Organisms* (Konings, W. N., Kaback, K. R. & Lolkema, J. S., eds) vol. 2, pp. 473–500, Elsevier Science B. V., Amsterdam.

Sturrock, S. S. & Collins, J. F. (1993). *MPsrch version 1.3*. Biocomputing Research Unit, University of Edinburgh, UK.

Sukharev, S. I., Blount, P., Martinac, B., Blattner, F. R. & Kung, C. (1994). A large-conductance mechanosensitive channel in *E. coli* encoded by *mscL* alone. *Nature*, **368**, 265–268.

Sukharev, S. I., Blount, P., Martinac, B., Guy, H. R. & Kung, C. (1996). MscL: A mechano-sensitive channel in *Escherichia coli*. In *Organellar Ion Channels and Transporters* (Clapham, D. E., ed.), pp. 133–141, The Rockefeller University Press, New York.

Tam, R. & Saier, M. H., Jr (1993). Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* **57**, 320–346.

Tate, C. G., Muiry, J. A. & Henderson, P. J. (1992). Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H+ transport protein in *Salmonella typhimurium* and *Escherichia coli*. *J. Biol. Chem.* **267**, 6923–6932.

Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M. & Venter, J. C. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*, **388**, 539–547.

Townsend, D. E., Esenwine, A. J., George, J., III, Bross, D., Maguire, M. E. & Smith, R. L. (1995). Cloning of the *mgtE* Mg2+ transporter from *Providencia stuartii* and the distribution of *mgtE* in Gram-negative and Gram-positive bacteria. *J. Bacteriol.* **177**, 5350–5354.

Tsai, K.-J., Hsu, C.-M. & Rosen, B. P. (1997). Efflux mechanisms of resistance to cadmium, arsenic and antimony in prokaryotes and eukaryotes. *Zool. Studies*, **36**, 1–16.

Varela, M. F. & Wilson, T. H. (1996). Molecular biology of the lactose carrier of *Escherichia coli*. *Biochim. Biophys. Acta*, **1276**, 21–34.

Vrljic, M., Sahm, H. & Eggeling, L. (1996). A new type of transporter with a new type of cellular function: L-lysine export from *Corynebacterium glutamicum*. *Mol. Microbiol.* **22**, 815–826.

Wolfram, L., Friedrich, B. & Eitinger, T. (1995). The *Alcaligenes eutrophus* protein HoxN mediates nickel transport in *Escherichia coli*. *J. Bacteriol.* **177**, 1840–1843.

Yan, R.-T. & Maloney, P. C. (1993). Identification of a residue in the translocation pathway of a membrane carrier. *Cell*, **75**, 37–44.

Yan, R.-T. & Maloney, P. C. (1995). Residues in the pathway through a membrane transporter. *Proc. Natl Acad. Sci. USA*, **92**, 5973–5976.

Young, R. & Bläsi, U. (1995). Holins: form and function in bacteriophage lysis. *FEMS Microbiol. Rev.* **17**, 191–205.